Structural, Item, and Test Generalizability of the Psychopathy Checklist–Revised to Offenders With Intellectual Disabilities

Catrin Morrissey,[1] David Cooke,[2] Christine Michie,[2] Clive Hollin,[3] Todd Hogue,[4] William R. Lindsay,[5] and John L. Taylor[6]

Abstract

The Psychopathy Checklist–Revised (PCL-R) is the most widely used measure of psychopathy in forensic clinical practice, but the generalizability of the measure to offenders with intellectual disabilities (ID) has not been clearly established. This study examined the structural equivalence and scalar equivalence of the PCL-R in a sample of 185 male offenders with ID in forensic mental health settings, as compared with a sample of 1,212 male prisoners without ID. Three models of the PCL-R's factor structure were evaluated with confirmatory factor analysis. The 3-factor hierarchical model of psychopathy was found to be a good fit to the ID PCL-R data, whereas neither the 4-factor model nor the traditional 2-factor model fitted. There were no cross-group differences in the factor structure, providing evidence of structural equivalence. However, item response theory analyses indicated metric differences in the ratings of psychopathy symptoms between the ID group and the comparison prisoner group. This finding has potential implications for the interpretation of PCL-R scores obtained with people with ID in forensic psychiatric settings.

Keywords

Psychopathy, PCL-R, intellectual disability, developmental disability, learning disability, intelligence, item response theory

The construct of psychopathy is currently conceptualized as a severe personality disorder characterized by a set of affective, interpersonal and behavioral features. These characteristics include deficient affective experience, a selfish, callous, and remorseless use of others, an impulsive and irresponsible lifestyle and, according to some theorists, antisocial behavior (Cleckley, 1941, 1976; Cooke & Michie, 1997; Hare, 2003; Hare & Neumann, 2005; Skeem & Cooke, in press). Assessment of the construct in forensic settings has focused on the Psychopathy Checklist–Revised (PCL-R; Hare, 1991, 2003), a measure based on expert ratings of 20 traits using information derived from semistructured interviews and file reviews (see Figure 1). As a measure, the PCL-R has demonstrated good internal consistency and interrater reliability across diverse forensic groups (Hare, 2003). The validity of the measure is also well established, and in particular PCL-R scores have been repeatedly shown to predict general criminal behavior and violent behavior at the group level (e.g., Douglas, Yeomans, & Boer, 2005; Guy, Edens, Anthony, & Douglas, 2005; Hemphill, Hare, & Wong, 1998; Walters, 2003). Although research has focused predominantly on adult male offenders and forensic psychiatric patients, more recent work has extended the application of the measure to female offenders (Forouzan & Cooke, 2005; Hare, 2003; Vitale, Smith, Brinkley, & Newman, 2002), and a modified version for youth has been developed (PCL:YV; Forth, Kosson, & Hare, 2003).

Various structural models for the PCL-R have been proposed, and there is much current debate regarding the strengths and weaknesses of the various models (e.g., Cooke, Michie, Hart, & Clark, 2004; Cooke, Michie, & Skeem, 2007; Hare & Neumann, 2005; Vitacco, 2007). Early exploratory factor analysis of large PCL-R data sets initially

[1]Rampton Hospital, Nottinghamshire, UK, and Institute of Mental Health, Nottingham, UK
[2]Glasgow Caledonian University, Glasgow, UK
[3]University of Leicester, Leicester, UK
[4]University of Lincoln, Lincoln, UK
[5]Tayside University, Tayside, UK and Castlebeck Care, Northumberland, UK
[6]Northumbria University, Newcastle, UK, and Northumberland, Tyne & Wear NHS Trust, Northumberland, UK

Corresponding author:
Catrin Morrissey, National High Secure Learning Disability Service, Rampton Hospital, Nottinghamshire Healthcare NHS Trust, Retford, Nottinghamshire, DN22 0PD, UK
Email: catrin.morrissey@nottshc.nhs.uk

Psychopathy		
Factor 1 Affective/Interpersonal	Factor 2 Social Deviance	No factor loading
Item 1. Glibness/superficial 2. Grandiose 4. Pathological lying 5. Conning /manipulative 6. Lack of remorse/guilt 7. Shallow affect 8. Callous/lack empathy 16. Failure to accept responsibility	Item 3. Need for stimulation 9. Parasitic lifestyle 10. Poor behavioral controls 12. Early behavioral problems 13. Lack realistic goals 14. Impulsivity 15. Irresponsibility 18. Juvenile delinquency 19. Revocation conditional release	Item 11. Promiscuous sexual behavior 17. Short-term marital relations 20. Criminal versatility

Figure 1. Psychopathy Checklist–Revised (PCL-R) traditional 2-factor, 20-item model, with item labels (Hare, 1991).

indicated a model with two correlated factors representing the Interpersonal/Affective elements of psychopathy and the Social Deviance elements, respectively (e.g., Hare, 1991; Harpur, Hakistan, & Hare, 1988; Kosson, Smith, & Newman, 1990). This factor structure had 8 items that loaded on Factor 1 and 9 items that loaded on Factor 2; 3 of the 20 items (Items 11, 17, and 19) did not load on either factor, but all items contributed to the overall construct of psychopathy. This is referred to as the traditional 2-factor model (see Figure 1). Cooke and Michie (2001), on the basis of both theory and confirmatory factor analysis (CFA) of several large data sets, developed a 3-factor hierarchical model of psychopathy using only 13 PCL-R items (see Figure 2). In this model, a coherent superordinate factor (Psychopathy) is underpinned by three correlated factors, which are labeled *Arrogant and Deceitful Interpersonal Style*, *Deficient Affective Experience* and *Impulsive*, and *Irresponsible Behavioral Style*. Below the level of specific factors are testlets that combine specific indicators to form higher-order facets within the hierarchy of personality features. Essentially, Cooke & Michie argued that this model is conceptually closer to the traditional personality-based model of psychopathy, and have since put the case (Cooke et al., 2004) that the antisocial behavior and relationship items it excludes are "downstream" manifestations of the core disorder.

In the second edition of the PCL-R manual (Hare, 2003), following analysis of further large data sets, an alternative hierarchical 2-factor, 4-facet model was described (see Hare, 2003, p. 78, Figures 7.1 and 7.4), which split each of the original two correlated factors into two subfactors or "facets" (Factor 1 into *Interpersonal* and *Affective*; Factor 2 into *Lifestyle* and *Antisocial*), all of which contributed to the overall superordinate construct of psychopathy (see Figure 3). Two items (Items 11 and 17) did not load on any facet, but did load on the superordinate factor. This is referred to as the 4-factor model (see Figure 3).[1] Hare (2003) states that the original 2-factor structure continues to be valid, although Item 20 (*Criminal versatility*) now loads on Factor 2. It should be noted that the three factors in Cooke and Michie's (2001) 3-factor model are identical in content to the first three factors of the 4-factor model, although they are labeled differently. Proponents of the 4-factor model consider the fourth factor, antisocial behavior, to represent a core aspect of the disorder (e.g., Hare & Neumann, 2005; Vitacco, 2007). For simplicity, the three models that have been described above are henceforth referred to as traditional 2-factor (Figure 1), 3-factor hierarchical (Figure 2), and a 4-factor hierarchical model (Figure 3), respectively.

Psychopathy and Offenders With Intellectual Disabilities

The term *intellectual disabilities* (ID)[2] denotes those individuals with significantly subaverage intellectual functioning and concurrent deficits in adaptive functioning, with an age of onset before 18 years. Although estimates of prevalence vary widely (Fazel, Xenitidis, & Powell, 2008), people with ID are likely to be overrepresented in various parts of the criminal justice system, and there has been recent formal recognition of the treatment needs of such individuals in U.K. settings (Department of Health, 2009). Although they form an important subgroup of all offenders, there had been little consideration given to the relevance of personality disorder generally, or psychopathy specifically, in forensic populations with ID. Some empirical research has now begun to explore assessment of these disorders in this group (Lindsay et al., 2006; Morrissey et al., 2005; Morrissey et al., 2007a; Morrissey, Mooney, Hogue, Lindsay, & Taylor, 2007b). Although this small body of work has indicated that severe personality disorders do occur in people with ID, it has been

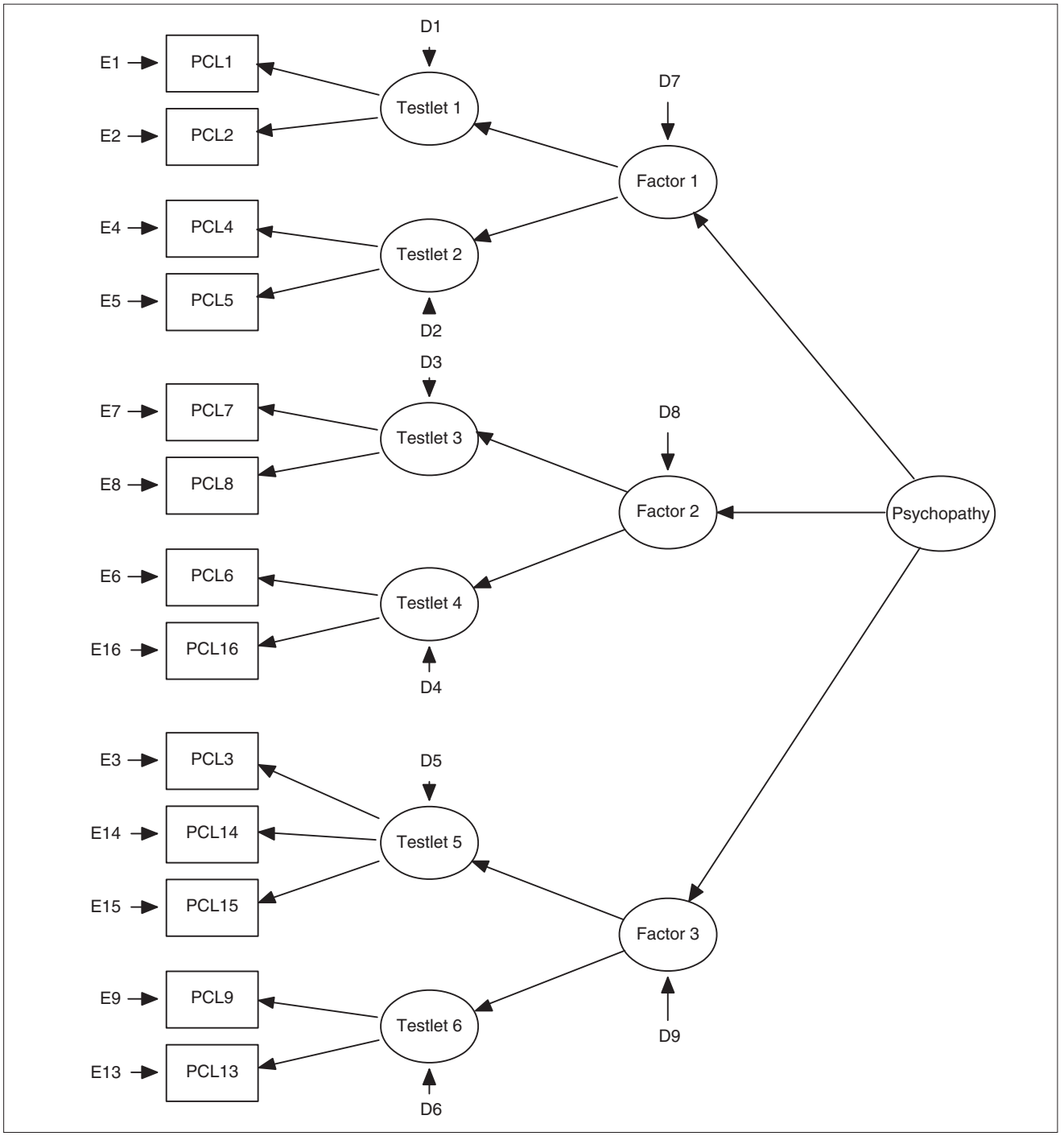

Figure 2. Psychopathy Checklist–Revised (PCL-R) hierarchical 3-factor, 13-item model with testlets (Cooke & Michie, 2001).

acknowledged that there remains considerable work to be done in respect of their valid and reliable measurement (see Alexander & Cooray, 2003). Such psychometric study is particularly important for measures of psychopathy because of its salience in various areas of forensic practice, including risk assessment, assessment of suitability for treatment, and use in capital cases (Edens, Petrila, & Buffington-Vollum, 2001).

We have previously reported on an initial study, which examined the generalizability of the PCL-R to 203 offenders

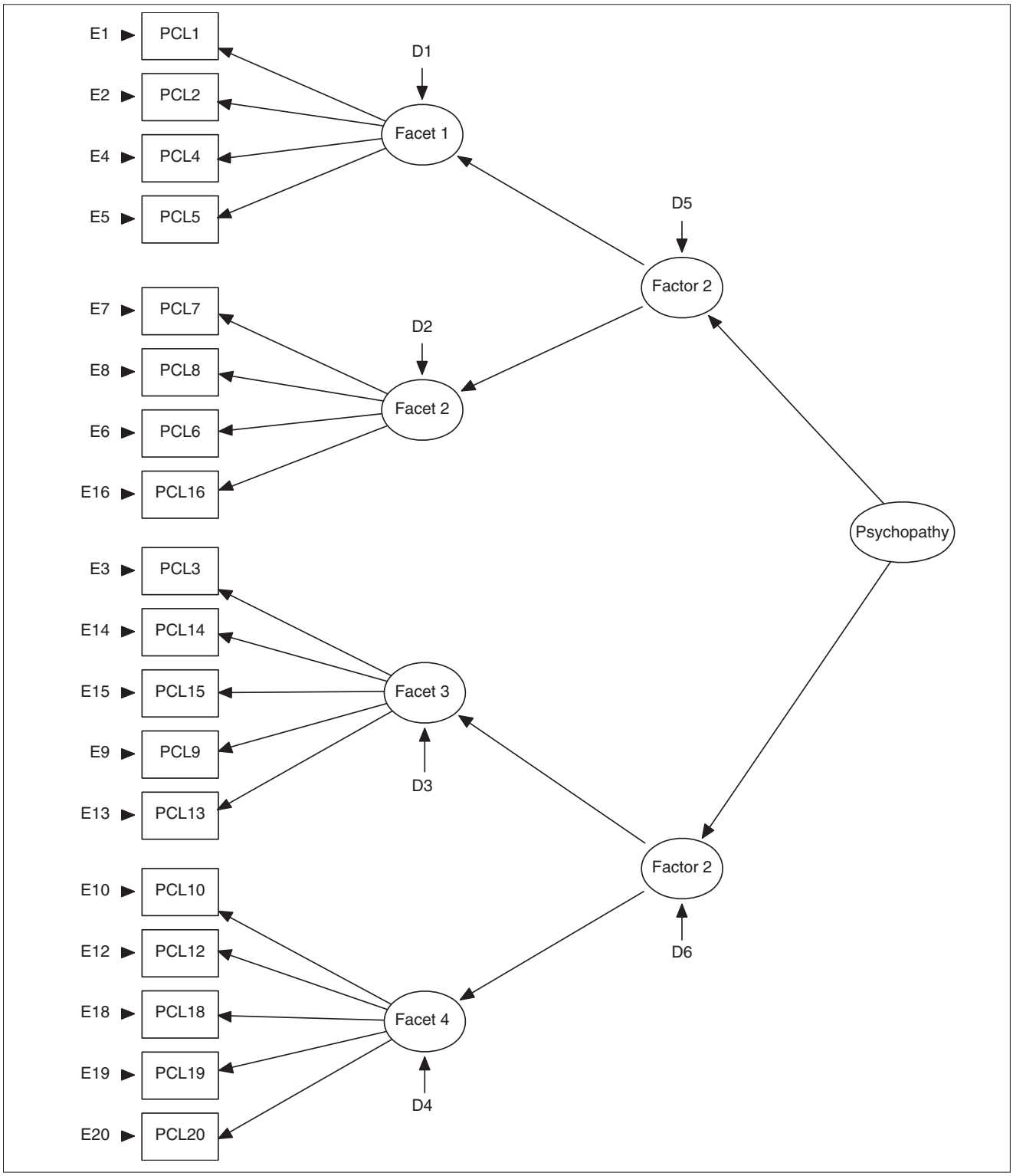

Figure 3. Psychopathy Checklist–Revised (PCL-R) hierarchical 4-factor, 18-item model (Hare, 2003).

with ID, and described its psychometric properties in terms of classical test theory (CTT; Morrissey et al., 2005). Using ratings derived from file review combined with interviews with clinicians who had detailed knowledge of the patient, it was found that the measure had adequate internal consistency and good interrater reliability. There was also some evidence for construct validity, with similar relationships with external criterion variables to those reported for

offender populations without ID being demonstrated. Furthermore, for the total sample, the mean PCL-R score, standard deviation, and proportion of individuals scoring above the conventional cut-off scores were noted to be very similar to those cited for U.K. male prisoners (e.g., Cooke, Michie, Hart, & Clark, 2005a; Hare, 2003). Notwithstanding these findings, the article also recognized that psychopathy may be expressed differently in people with ID, and that there were both methodological and conceptual problems with using the PCL-R with this population. It was also recommended that further analyses were necessary to adequately establish the generalizability of the measure to people with impaired intelligence and social functioning.

In the psychometric literature, it has been argued that the statistics commonly used to describe the distribution and structure of scores on a measure do not provide an adequate test of whether total scores, individual items, and factor structure, operate equivalently across different groups of people (e.g., Cooke, Kosson, & Michie, 2001; Cooke, Michie, Hart, & Clark, 2005b; Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991). To examine such issues, it has been recommended that the focus should not be on manifest variables such as test scores, which can be biased, but rather on the latent variable that underlies them—in the current context, the *latent trait* of psychopathy. In previous studies exploring the cross-group generalizability of the PCL-R, Cooke and colleagues (Cooke & Michie, 1999; Cooke et al., 2001; Cooke et al., 2005a, 2005b) have noted that both *structural equivalence* and *scalar equivalence* need to be established. In terms of structural equivalence, it first needs to be determined whether the factor structure of this latent trait is the same across groups. This can be assessed by the application of CFA, in which a specifically hypothesized set of latent variables is fitted to a covariance matrix. CFA can determine if the number of factors present, the variable to factor relationships and factor to factor relationships conform to what is expected on the basis of a preestablished model, and can also assess whether there are significant group differences in factor structure. To determine scalar functioning, it needs to be established whether the ratings on individual PCL-R items and the scores obtained on the whole scale (or factor scales) have the same relationship to the latent trait of psychopathy in the different groups of interest. These questions can be explored using item response theory (IRT) methods (Embretson & Reise, 2000). IRT is a model-based version of test theory that applies rules of measurement that are fundamentally different to the "old" rules of CTT. The methods focus on underlying latent traits (usually represented as θ, and standardized to have a mean of 0 and a standard deviation of 1), and can provide useful information about the properties of items and whether or not numerically equivalent levels of total scores on a test represent the same level of the latent trait in two (or more) groups.

With the aim of exploring structural and scalar equivalence respectively, this article therefore describes CFA and IRT analysis of PCL-R data obtained from a sample of U.K. offenders with ID, and a large previously studied U.K. prison comparison sample without such disabilities identified. These procedures have been applied together in validating the PCL-R across different populations (e.g., gender, ethnicity, cultural groups) in a number of studies (e.g., Bolt, Hare, Vitale, & Newman, 2004; Cooke & Michie, 1999; Cooke et al., 2001; Cooke et al., 2005a, 2005b) but not, until now, for different levels of intellectual and adaptive ability. The basis of these techniques and IRT models and methods in particular, are described in detail in these previous studies.

Method

Participants

The participants were 185 males (full scale IQ <75, mean IQ = 66)[3] who were sampled from three U.K. National Health Service forensic psychiatric services catering specifically for offenders with ID (a high-security hospital, a medium-to-low security hospital, and a community forensic service). All participants had a clinical diagnosis of intellectual disability, and were deemed to have significant deficits in intellectual functioning, with concurrent deficits in adaptive functioning, with an age of onset before 18 years. These individuals were included in a wider study of risk and personality disorder, and the sample is described in more detail elsewhere (see Hogue et al., 2006). The comparison U.K. prison sample comprised a previously studied combined sample of 1,212 adult male offenders—608 adult male offenders from seven prisons in England and Wales, a representative sample of 246 offenders from the Scottish Prison Service, a separate stratified random sample of 253 offenders from Scotland's largest prison, and a sample of 105 incarcerated offenders who participated in a study of early childhood experiences. The prison sample, and related data analyses, is described further in Cooke et al. (2005a).

Comparative demographic information was not available for all the samples. The sample did differ in terms of age, with the ID sample being older than the prisoners ($t = -10.47$, $df = 967$, $p < .001$).[4]

Measure

The PCL-R consists of 20 items (see Figure 1). Each item is scored on a 3-point scale (0 = *absent*, 1 = *maybe/in some respects*, or 2 = *present*), indicating the degree to which the item applies to the individual. Higher scores are indicative of a greater number and/or severity of psychopathic characteristics. Psychometric properties of the measure are given in Hare (2003). The psychometric properties of the PCL-R with the current ID sample are described further in Morrissey et al. (2005).

Procedure

For the ID sample, the PCL-R was scored from file review combined with interviews with clinicians (psychiatrists and psychologists with clinical responsibility for the patient). All raters were trained in standard administration of the PCL-R, and were informed by supplementary guidelines for using the PCL-R in people with intellectual disabilities developed by the first author for the purposes of the study (Morrissey, 2005).[5] These guidelines were intended to increase interrater reliability on coding difficulties typical in ID assessments, but adhered to the standard PCL-R manual (Hare, 2003) in every respect. Mean PCL-R total score was 15.4 (SD = 8.05) for the prisoner sample and 16.1 (SD = 7.23) for the ID sample. For the ID sample interrater reliability was assessed in 45 cases where there were two independent ratings (total score intraclass correlation coefficients, $ICC_{A,1}$ [absolute agreement for a single rater] = .89). In all the prison samples, the standard method of administration (file review and patient interview) was employed. For the Scottish prisoner sample, total score $ICC_{A,1}$ = .86 (Cooke & Michie, 2009).

Data Analyses

Confirmatory factor analysis. CFA was performed using EQS (Bentler & Wu, 1995) specifying categorical variables, to test the fit of the three main hypothesized factor structures for the PCL-R: the traditional 2-factor model, the 3-factor hierarchical model with testlets, and the 4-factor model (Figures 1, 2, and 3). CFA was first performed on the PCL-R data from the ID group alone. A multigroup CFA was then performed on the ID and Prisoner group simultaneously. Cases with missing data were deleted listwise from these analyses. Maximum likelihood estimation with robust fit statistics and standard errors was used. The correlations were polychorics. Recommendations in the electronic help manual for the EQS 6 software suggests that this estimation approach is the best EQS approach for data of this type.

Item response theory. IRT models estimate the relationship between item or test scores and the latent trait (θ) that underlies them. Essentially the trait level for the individual is estimated from the pattern of their item scores taking into account the characteristics of individual items. Item characteristic curves (ICCs) index the association between the probability of an item score θ, and test characteristic curves (TCCs) index the association between the probability of total scores and θ. The slopes of the ICCs or TCCs reflect discriminating power, that is the extent to which item or test scores discriminate between those high and low on the latent trait. The inflexion point of ICCs and TCCs reflects the extremity or difficulty of item and test scores. IRT methods can also be used to detect differences in item functioning or test functioning across groups. Differential item functioning occurs when an item is more discriminating, or is more difficult or extreme, in one group as compared with another, and differential test functioning occurs when the total scores on a test are more discriminating or more extreme in one group than another.

Different IRT models use different mathematical functions, based on diverse sets of assumptions. For the purposes of the current analysis, the Graded Response Model (Samejima, 1969) was used (following Cooke et al., 2001; Cooke et al., 2005a; Cooke & Michie, 1997), as the assumptions of the model are suitable for the ordinal/categorical data obtained from PCL-R items. The probability of the three response options (0, 1, and 2) for a PCL-R item in relation to the level of the latent trait (psychopathy) can be expressed by probability curves. The shape and position of the curves can be summarized by the values of three parameters a, b_1, and b_2. The a parameter is an index of slope (i.e., discrimination), higher values indicating better discrimination, and the b parameters are indexes of difficulty or extremity, higher values indicating a higher threshold.

The IRT analysis was conducted using Multilog VII (Thissen, 2003).

Results

Initial CFA Analysis: ID Sample

Factor structures reflecting the three models (traditional 2-factor, 3-factor [with testlets], and 4-factor) were specified and modeled for each proposed factor solution, and their fit to the PCL-R data from the ID sample was analyzed. The quality of "fit" to each model to the ID PCL-R data was estimated using multiple measures of fit, as there is no one fit index that is widely accepted as the "gold standard" given that all measures have limitations (Cooke & Michie, 2001; Kline, 1998). The indices used were (a) the Satorra–Bentler (S-B) scaled chi-square (Satorra & Bentler, 2001) (b) the nonnormed fit index (NNFI) (c) the comparative fit index (CFI), and (d) the root mean square error of approximation (RMSEA). The criteria for an adequate fit were defined as CFI and NNFI ≥.95 and RMSEA ≤.05.

The fit for the 3-factor hierarchical model was good, S-Bχ^2(56, N = 174) = 80.7, nonsignificant (ns); NNFI = .95; CFI = .96; RMSEA = .05. The 4-factor hierarchical model did not meet the criteria for fit, S-Bχ^2(130, N = 155) = 224.7, p < .001; NNFI = .87, CFI = .89; RMSEA = .07. The traditional 2-factor model demonstrated a poor fit, S-Bχ^2(118, N = 155) = 272.6, p < .001; NNFI = .77; CFI = .80; RMSEA = .09. It may be concluded that the 3-factor hierarchical structural model fits the ID PCL-R data well, whereas neither the 4-factor hierarchical model nor the traditional 2-factor model fit the data.

Table 1. Fitted Item Response Theory Parameters for All 20 Items: Unconstrained Baseline Model

Item		Intellectual Disabilities			Prisoner		
		a	b_1	b_2	*a*	b_1	b_2
1	Glibness/superficial charm	1.1	0.6	2.6	1.0	0.9	2.9
2	Grandiose sense of self-worth	1.4	0.4	2.0	1.2	0.2	1.6
3	Need for stimulation	0.7	–0.2	2.8	1.4	–0.9	0.8
4	Pathological lying	1.3	0.2	1.8	1.1	0.2	1.8
5	Conning/manipulative	1.0	–0.5	1.6	1.3	–0.6	1.2
6	Lack of remorse or guilt	3.2	–1.0	0.1	1.6	–1.3	0.0
7	Shallow affect	1.9	0.4	3.1	1.5	–0.4	1.1
8	Callous/lack of empathy	2.6	–1.0	0.4	1.8	–0.4	1.0
9	Parasitic lifestyle	0.9	1.1	3.0	1.1	–1.2	1.1
10	Poor behavioral controls	0.7	–1.9	0.5	1.2	–1.0	0.5
11	Promiscuous sexual behavior	0.3	–2.0	3.1	0.8	–0.7	0.9
12	Early behavior problems	0.6	–2.4	–0.3	1.4	–0.4	0.4
13	Lack of long-term goals	1.3	0.0	1.5	1.0	–0.8	0.7
14	Impulsivity	0.7	–1.8	1.0	1.1	–1.0	0.8
15	Irresponsibility	0.6	–0.1	3.0	1.1	–1.5	0.9
16	Failure to accept responsibility	2.2	–1.0	0.3	1.0	–1.6	0.8
17	Short-term marital relationships	0.9	3.2	4.6	0.7	0.7	2.0
18	Juvenile delinquency	0.5	–0.3	1.6	1.1	–1.3	0.0
19	Revocation of conditional release	0.6	0.5	1.7	0.8	–1.0	0.2
20	Criminal versatility	0.7	1.2	3.5	1.0	–1.2	0.1

The U.K. Prison sample had previously been found to fit the 3-factor model, but not the 2-factor or the 4-factor model (see Cooke et al., 2005a; Cooke et al., 2007).

Multigroup CFA Analysis

Multigroup CFA analyses were then conducted using EQS to establish more specifically whether the structure of the PCL-R in the ID sample was the same as that observed in the prison comparison group. This process involved fitting the 3-factor model simultaneously to data from the ID sample and the Prison sample. First, an unconstrained baseline model was fitted (i.e., all the parameters were allowed to take different values for the two samples), and this model was a good fit, S-Bχ^2(112, N = 1,284) = 268.7, $p <$ .001; NNFI = .96; CFI = .97; RMSEA = .05. A model in which the factor loadings of the PCL-R items on the three factors were constrained to be equal for the ID and Prison groups was then analyzed. The model was also found to fit the data adequately, S-Bχ^2(125, N = 1,284) = 397.5, $p <$ .001; NNFI = .93; CFI = .95; RMSEA = .06, and the fit in comparison to the unconstrained model was not significantly degraded (as indicated by a scaled difference chi-square test [Satorra & Bentler, 2001]: ΔS-Bχ^2(13, N = 1,284) = 19.9, ns. It can therefore be concluded that the factor structure across the two groups was invariant.

IRT Analysis

Although the CFA analysis had indicated that only the 3-factor model fitted the data, because most researchers and clinicians use all 20 items, we examined the IRT parameters for both the 3-factor 13-item model and for all 20 items included in the 4-factor model.[6]

Comparison of item characteristic curves and item parameters. The IRT analysis has a number of stages. First the two samples are considered separately allowing the latent trait and item parameters to vary (unconstrained baseline model), then the parameters for the two groups are compared. If there is a difference in the parameters, procedures are carried out to put the groups on a common latent trait metric so that the scores and TCCs can be compared.

Table 1 shows the fitted item parameters (a, b_1, b_2) for both the ID sample and the Prison sample for the 20 items (unconstrained baseline model). (Results for the items in the 13-item test were similar, and are therefore not displayed). The parameters were first examined for the ID group; the discrimination, a (i.e., slope) parameters for nine of the items are .7 or less, values which indicate poor discrimination. All five items in the antisocial facet (Factor 4 of Hare's 4-factor model) are less than .7, and for the lifestyle factor (Factor 3), three of the five items have a values of .7 and less. In contrast, discrimination parameters for

Table 2. Fitted Item Response Theory Parameters for 13- and 20-Item Models: Constrained Model Anchored on Factor 1

	13-Item Model						20-Item Model					
	Intellectual Disabilities			Prisoner			Intellectual Disabilities			Prisoner		
Item	a	b1	b2	a	b1	b2	a	b1	b2	a	b1	b2
1. Glibness/superficial	**1.2**	**0.6**	**2.4**	**1.2**	**0.6**	**2.4**	**1.0**	**0.8**	**2.8**	**1.0**	**0.8**	**2.8**
2. Grandiose	**1.5**	**0.2**	**1.4**	**1.5**	**0.2**	**1.4**	**1.2**	**0.3**	**1.7**	**1.2**	**0.3**	**1.7**
3. Need for stimulation	0.6	–0.3	3.0	1.1	–0.9	0.9	0.7	–0.3	2.7	1.4	–0.9	0.8
4. Pathological lying	**1.2**	**0.1**	**1.7**	**1.2**	**0.1**	**1.7**	**1.2**	**0.2**	**1.8**	**1.2**	**0.2**	**1.8**
5. Conning/manipulative	**1.2**	**–0.6**	**1.3**	**1.2**	**–0.6**	**1.3**	**1.2**	**–0.6**	**1.2**	**1.2**	**–0.6**	**1.2**
6. Lack of remorse/guilt	3.3	–1.1	0.0	1.9	–1.3	0.0	3.2	–1.1	0.0	1.6	–1.3	0.0
7. Shallow affect	1.0	0.3	2.8	1.7	–0.4	1.0	0.9	0.3	3.0	1.5	–0.4	1.1
8. Callous/lack of empathy	2.6	–1.0	0.3	2.1	–0.4	0.9	2.6	–1.0	0.4	1.8	–0.4	1.0
9. Parasitic lifestyle	1.0	1.0	2.8	0.9	–1.3	1.4	1.0	1.0	2.9	1.1	–1.2	1.2
10. Poor behavioral control							0.7	–2.0	0.5	1.2	–1.0	0.5
11. Promiscuous sexual							0.3	–2.0	3.0	0.8	–0.6	0.9
12. Early problems							0.6	–2.4	–0.4	1.4	–0.4	0.4
13. Lack of long-term goals	1.3	–0.1	1.4	1.0	–0.8	0.7	1.2	0.0	1.5	1.0	–0.8	0.7
14. Impulsivity	0.6	–2.1	1.1	0.9	–1.1	1.0	0.7	–1.8	0.9	1.1	–1.0	0.8
15. Irresponsibility	0.6	–0.2	3.0	0.9	–1.6	1.0	0.6	–0.2	3.0	1.1	–1.4	0.9
16. Failure to accept	2.4	–1.1	0.2	1.2	–1.5	0.6	2.3	–1.1	0.3	1.0	–1.6	0.8
17. Short-term marital							0.9	3.2	4.6	0.7	0.7	2.0
18. Juvenile delinquency							0.5	–0.3	1.6	1.1	–1.3	0.0
19. Revocation conditional							0.6	0.4	1.6	0.8	–1.0	0.2
20. Criminal versatility							0.7	1.1	3.4	1.0	–1.2	0.1

Note: Entries in boldface indicate items constrained to be equal across groups.

items in Facet 1 (interpersonal) and Facet 2 (affective) are high, indicating good discriminative power. In general, it was evident that the *b* (i.e., threshold) parameters for items in Factor 1 (interpersonal) and Factor 3 (lifestyle; in both the 3- and 4-factor models) were higher than those for most items in Factor 2, indicating that these interpersonal and lifestyle manifestations symptoms only become apparent at higher levels of the disorder in the ID sample. Conversely most of the Factor 2 (affective) symptoms are present even at low levels of the disorder.

The parameters for the two groups (ID and prisoners) were then compared statistically. Equality of slopes (*a* parameters) across the two groups would indicate that items have similar relevance for defining the underlying psychopathy trait in each group. The equivalence of parameters across groups can be determined by comparing the goodness of fit of a model where the parameters are constrained to be equal with a goodness of fit of an unconstrained model using a generalized likelihood ratio test. If there are no significant differences between the models there is no evidence of differences in the item parameters. For the 13-item test, testing for whether all parameters were equal indicated that this was not the case, $\Delta\chi^2(39, N = 1{,}397) = 396.2$, $p < .001$. Testing for whether *a* (slope) parameters were equal across the groups also indicated that this was not the case, $\Delta\chi^2(13, N = 1{,}397) = 44.0$, $p < .001$. It can be concluded that there are differences in the item discrimination parameters between the two groups (ID and prisoners). However, the parameters for the four Facet 1 items (Items 1, 2, 4, and 5) in the ID group were noted to be very similar in both groups, and testing indicated that it was possible to constrain these parameters to be equal without degrading the fit of the model, $\Delta\chi^2(12, N = 1{,}397) = 14$, ns. Testing was repeated for the 20-item model and similar results were obtained.

The above findings indicate differential item functioning (DIF) across the two groups (ID and prisoners), which means that estimates of the underlying trait were not directly comparable. Direct comparison of the two groups therefore necessitated additional procedures.

One advantage of IRT methods is that it is not essential for all items to have similar parameters to ensure that a common metric underpins scores in different groups. So, even if items in a scale behave differently across samples, as long as there is a core group of items that behave in the same way across groups, these invariant items can act as "anchors" to establish a common metric on which to measure the psychopathy trait (θ) in both groups. This common metric means that, although they may have different PCL-R scores, an ID individual whose score on θ is 1.0 will be equivalent in terms of degree of psychopathy to a prisoner

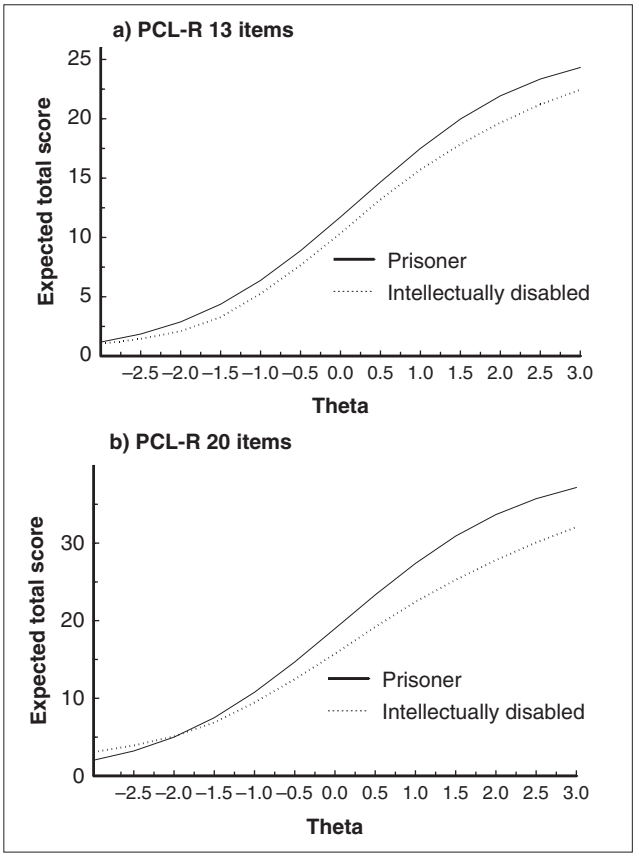

Figure 4. Test characteristic curves for (A) 13-item model and (B) 20-item model: ID sample and Prisoner sample.

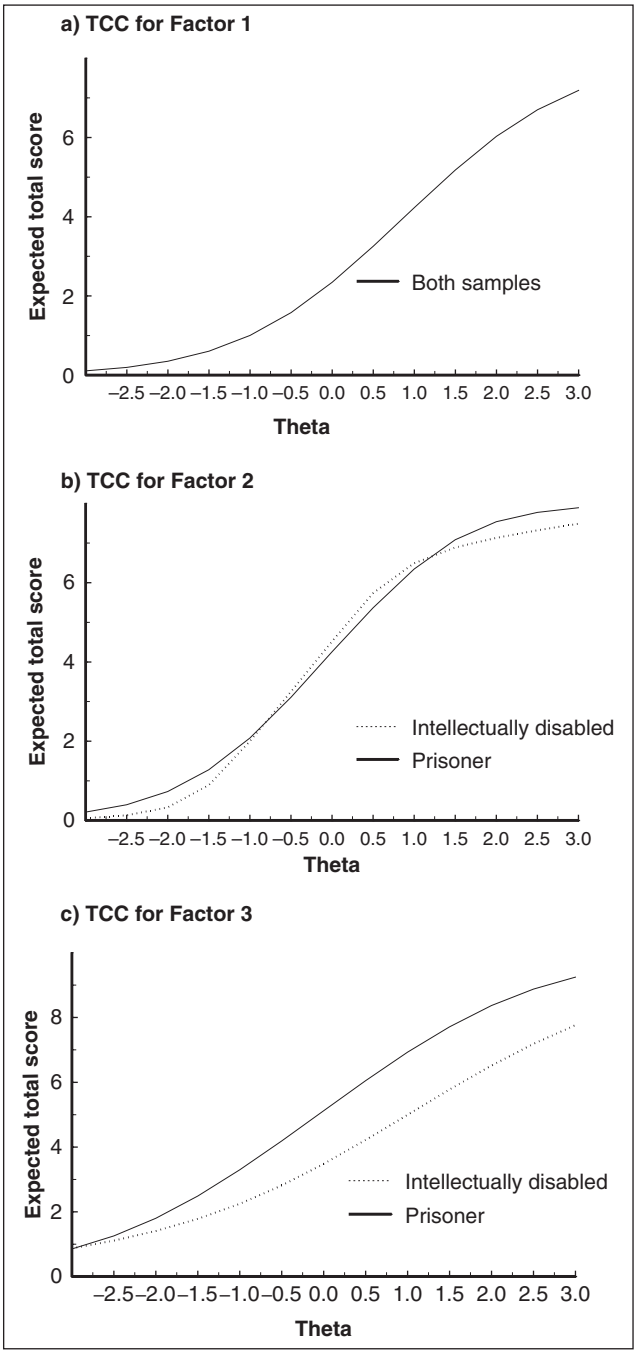

Figure 5. Test characteristics curves for 13-item model (A) Factor 1, (B) Factor 2, and (C) Factor 3: ID sample and Prisoner sample.

whose score on θ is 1.0. As it had been established that there was no difference in the parameters of the four Factor 1 items across the two groups, these four items were therefore used as anchors; this method is described further in Cooke and Michie (1999). Parameters after anchoring (the constrained model) for both samples are displayed in Table 2 for the 20-item model and the 13-item model. Visual examination of Table 2 suggests DIF for most of the Factor 3 (lifestyle) items (in both models), and for most of the Facet 4 (antisocial) items (20-item model). Broadly speaking, these items show poorer discrimination and higher thresholds in the ID sample, as compared with the prison sample.

Comparison of test characteristic curves. Differential test functioning (i.e., bias at the whole test level) was then assessed by examining the TCCs. TCCs were considered for the 13-item total score (Total 13; Figure 4A), as well as for the conventional 20-item total score (Total 20; Figure 4B). Visual inspection of the Total 13 and Total 20 curves indicates that the slope is less steep for the ID sample as compared with the prison sample, which suggests that the test as a whole is less discriminating for the ID sample. Also, for equivalent levels of psychopathy (θ), scores on both the Total 13 and on the Total 20 are lower in the ID sample than in the U.K. Prison sample, particularly at higher levels of the latent trait.

A numerical index of differential test functioning, the root differential test function (rDTF; Raju, Linden, & Fleer, 1995) can be calculated. The rDTF index expresses the

average differences in TCCs in the metric of the test in raw score units. For the 13-item total, the rDTF was 1.4 points (out of a maximum score of 26), and on the 20 item total the rDTF was 2.8 points (maximum score 40; both $p < .001$). The value of the rDTF indicates the number of points which, on average, needs to be added to the ID PCL-R score, to equate to the Prisoner PCL-R score (so for example if an ID score was 20, it would equate to a Prisoner score of 22.8). However it should be noted that this figure is an average; visual examination of the TCCs indicates that differences are larger at higher levels of the underlying trait, particularly for the 20-item scale. This means that for the ID group the scale particularly underestimates the latent trait at higher levels of the disorder. For example, if a cutoff of 30 is chosen to denote a high psychopathy score, then 2% of the ID sample would be labeled as such. However, the TCCs suggest that a score of 30 in the prisoner sample is equivalent to a score of 25 in the ID sample; if 25 were used instead as the cut-off score, 11% would be regarded as having a high level of psychopathic traits.

To obtain a clearer picture of differences across groups, the TCCs for each of the three lower-order factors within the 3-factor model only were examined (see Figures 5A, 5B, and 5C). It should be noted that as Factor 1 parameters (*Arrogant and deceitful interpersonal style*: Figure 5A) were anchored with each other, the curves are identical. Nevertheless, as the parameters prior to anchoring were very similar, it can be assumed there was not differential test functioning on Factor 1. It is evident from inspection of the TCC in Figure 5C that Factor 3 in particular (*Impulsive and irresponsible behavioral style*) is markedly underestimating psychopathy at all levels of the trait in the ID sample, as compared with the Prison sample. For Factor 2 (*Deficient affective experience:* Figure 5B), the curves for the two groups are more similar to each other, although psychopathy is still underestimated for the ID group at high and low levels of the trait. Examination of the slopes at the point of inflection indicates that, of the three factors, Factor 3 is clearly the least discriminating for the ID group. In terms of threshold, it is evident that very high levels of psychopathy are necessary before a high score on Factor 3 is achieved in the ID sample.

The rDTF index was calculated for Factor 2 (0.25 for a maximum score of 8; $p > .05$) and Factor 3 (1.44 for a maximum score of 10; $p < .001$), confirming that differential test functioning is most marked for Factor 3. It can be concluded that cross-group differences between the ID sample and the prison sample are therefore largest for the features of psychopathy reflected in an irresponsible and impulsive behavioral style.

Discussion

Our findings provide mixed evidence for equivalence of functioning of the PCL-R between those with and without intellectual disability. The CFA analyses found relatively good evidence for what has previously been referred to as "syndromal equivalence" (Cooke et al., 2005b) between the ID sample and a comparative U.K. prison sample. In the current CFA, the 3-factor hierarchical model (Cooke & Michie, 2001) fitted the ID PCL-R data well, and importantly no cross-group differences were detected in the factor structure. This suggests that the same characteristics relate together in the same way to make up the construct of psychopathy in the two populations. In other words, the same construct or latent trait is being assessed in the lower ability group as in the higher ability group. These findings indicate that the overarching disorder defined by the symptoms is a coherent syndrome in the ID group, which implies that psychopathy as a construct can be generalized to offenders with intellectual disabilities.

However, when IRT models are used to compare ID and Prisoner samples, there was evidence of both differential item functioning and differential test functioning on the PCL-R. First, there were significant differences in the discrimination parameters obtained for the two groups. This suggests that different characteristics discriminate between low and high scorers in those with and without ID. In the ID group, many items were found to discriminate poorly; in particular, several Factor 3 items (in both the 3- and 4-factor models) and Factor 4 of the 4-factor model (antisocial behavior items), as well as Item 11 (*Promiscuous sexual behavior*) and Item 17 (*Many marital relationships*). These items therefore have little value from a psychometric perspective. Although the approaches and assumptions are very different in IRT and classical test theory, these findings support our previous findings of low alpha values for Factors 3 and 4, and low item–total correlations for Items 11 and 17 in earlier reliability analyses (Morrissey et al., 2005). In contrast, discrimination parameters for items in Factor 1 and to a lesser extent for Factor 2 were uniformly high, and more similar to those observed in the Prison group.

Of more concern, at the whole test level, the IRT analyses indicated that there is evidence of "metric differences" in the ratings of psychopathy symptoms between the two groups. The scores obtained on the PCL-R in the ID and Prison groups (for both the 13-item and 20-item total) cannot therefore be considered equivalent. Overall, PCL-R scores were lower in the ID sample than in the U.K. Prison sample *given equal levels of the latent trait of psychopathy*. In practice, this means that PCL-R scores in an ID group are typically underestimating the "true" level of psychopathy (although this was true to a lesser extent for the 13-item model). We conclude that caution is necessary when interpreting total scores against comparative data, and particularly if applying the commonly used cutoff scores for clinical decision making. Clearly, any adjustment of PCL-R cutoff scores for people with ID would be premature and would have serious implications. For example, adjustment to a cutoff for a high

score from 30 to 25 based on the study findings would increase the proportion considered to have high levels of psychopathy from 2% to 11%.

When TCCs at the factor level (for the 3-factor model) were considered, differential test functioning was most evident for Factor 3 (*Impulsive and irresponsible behavioral style*), with evidence that ID scores underestimate psychopathy at all levels of the trait in comparison to the Prison group. It is notable that this factor included a number of items that have been previously identified as presenting scoring difficulties in an ID population (Morrissey et al., 2005). From the perspectives of both research and clinical experience with intellectual disability populations, it is predictable that both the lower level of cognitive and social skills and the typical constraints on the life of a person with ID would suppress the ease with which some of the "lifestyle" and "behavioral style" characteristics indicative of psychopathy are expressed. People with ID in community settings in the United Kingdom generally live in staffed accommodation, and until the advent of community care policies would have commonly lived in institutions. Employment opportunities are few, for other than the most able, and individuals are largely dependent on others or the state from both a practical and financial perspective. Opportunities for social and intimate relationships are also very restricted (Murphy, 1992). As a consequence, compared with intellectually able people, there are fewer opportunities to demonstrate parasitic and irresponsible traits (as defined by the PCL-R). Furthermore, because employment and normal family, social, and intimate relationships are not generally accessible to individuals with ID, there are inevitably a restricted range of domains for which evidence is available to score several other PCL-R items.

Although not included in Cooke and Michie's (2001) 3-factor model of psychopathy, the items relating to sexual and marital relationships (Items 11 and 17) had either poor discrimination or very high threshold values, in the item-level IRT analyses. Once again, this is unsurprising given the low probability of achieving such relationships for an ID population. It is further evident from the item-level analyses that some of the antisocial features of psychopathy appear to be suppressed in a similar way by the limitations associated with intellectual disability, and by what may be referred to as the "culture" of ID. For instance, lower levels of cognitive functioning reduce the capacity for certain types of more sophisticated offending behavior or behaviors that involve planning (such as fraud, absconding, and escape); this has a potential impact on a number of the PCL-R antisocial behavior items (e.g., *Criminal versatility*, *Revocation of conditional release*). Similarly, it can be assumed that restricted opportunities for offending are afforded by the living environments that are typical for people with ID (e.g., staffed supported accommodation or long-stay institutions). Crucially, it is also widely recognized that, for a number of reasons, people with ID are less likely to be processed through the criminal justice system for "offending like" behavior than those without such disabilities (e.g., Holland, Clare, & Mukhopadhyay, 2002; Lyall, Holland, & Collins, 1995; McBrien & Murphy, 2006). It follows that any items reliant on criminal history data will potentially underrepresent the antisocial behavior of people with intellectual limitations. We would posit that the current data from people with intellectual disabilities supports the contention that some of the lifestyle, relationship, and the antisocial manifestations of psychopathy are "characteristic adaptations" (Lilienfield, 1994; McCrae & Costa, 2003), which can be influenced by the environment, rather than "basic tendencies" of the psychopathy disorder.

In contrast, the findings indicate that Factor 1 (*Arrogant and deceitful interpersonal style*) and, to a lesser extent, Factor 2 (*Deficient affective experience*) functioned more similarly across the ID and non-ID groups, suggesting that the interpersonal and affective aspects of the disorder are more stable across the range of intellectual abilities. Because these two factors are considered by some to be most central to psychopathy (Cleckley, 1941, 1976; Cooke & Michie, 1997; Skeem & Cooke, in press), it could be argued that these are the elements on which focus should be placed when making clinical assessments of psychopathy in people with intellectual abilities and adaptive skills below the normal range. Furthermore, because the interpersonal and affective features of psychopathy have also been found to be the most stable when different ethnic groups were compared on the PCL-R (Cooke et al., 2001; Cooke et al., 2005b), and when females were compared with males (Bolt et al., 2004), the current data could be seen to provide indirect support for the claim that the affective and interpersonal features of psychopathy are indeed at the "core" of the disorder.

There are a number of important weaknesses to the comparative study which need to be noted. First, the ID sample size, although sufficient, is relatively small for both IRT analysis and CFA analysis (Cooke et al., 2007; Kline, 1998), and therefore the analysis has modest power. Second, there is the possibility that some of the differences found between the groups in the IRT analysis could be attributable to the different method of administration of the PCL-R (i.e., patient interview vs. informant interview). However, previous research suggests that this would be most likely to result in differences in Factors 1 and 2 (i.e., the interpersonal and affective presentation; see Bolt et al., 2004), which were in fact found to be broadly comparable. The differing methods of administration would therefore be very unlikely to explain the much greater difference found on Factor 3 and Factor 4, which depend less on a face-to-face interview. Third, the ID sample was not matched to the prison sample on key demographic and clinical variables. In particular, the ID sample was from forensic psychiatric settings where a proportion of patients would

have other comorbid mental disorders. Such comorbidity may not have been a significant factor in the prison samples, and again it is possible that the differences observed between the groups reflect this differing aspect of clinical presentation. Fourth, because IQ was not available for the prison sample, the extent of the overlap in intellectual ability between the two groups cannot be determined. Although persons with clearly diagnosable ID (i.e., deficits in adaptive functioning as well as subaverage IQ) are estimated to form only 0.5% to 1.5% of those in prisons (Fazel et al., 2008), some degree of overlap in IQ range is possible between the two groups under study. Nevertheless, it is important to note that any such overlap would actually mitigate *against* finding a difference between the two groups.

In conclusion, although we found there were metric differences in the measurement of psychopathy between ID patients and prisoners, we acknowledge that there cannot be any certainty that the differences observed in this study are wholly attributable to the differences in intellectual ability (and associated adaptive functioning) between the two samples. The findings reported in this study therefore require replication, preferably with an even larger sample of ID participants, which is matched more closely to a comparable non-ID group. Notwithstanding these limitations, the current findings do suggest that caution should be applied when using total PCL-R scale scores (whether from the 13- or 20-item test) for clinical purposes in ID forensic populations, particularly if scores in "normal" prison-based populations are used as a benchmark. We therefore favor using the measure to provide support for clinical formulation in people with ID, as opposed to focusing on total scores to determine degree of psychopathic features. The findings also lead us to hypothesize that the PCL:SV (screening version: Hart, Cox, & Hare, 1995) is potentially a more appropriate measure of psychopathy than the PCL-R for use with people with ID, as it excludes the relationship items and its antisocial behavior items are more broadly based than those in the PCL-R. It is noted that the PCL:SV has recently been found to have reasonably good predictive validity in forensic patients with ID (Gray, Fitzgerald, Taylor, MacCulloch, & Snowden, 2007), although it would seem important that IRT analyses should similarly be applied to PCL:SV data in order to establish scalar equivalence across populations. A final observation relates to the fact that it is the more behavioral elements of the PCL-R (i.e., Factors 3 and 4 of the 3- and 4-factor models, or Factor 2 of the 2-factor model) that are generally found to be the better predictors of antisocial behavioral outcomes (e.g., Walters, 2003). The differential functioning of these aspects of the measure in people with ID found in the current study should therefore be considered in any future predictive validity studies using the PCL-R with this important subgroup of offenders.

Declaration of Conflicting Interests

The authors declared no conflicts of interests with respect to the authorship and/or the publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: The DSPD Programme, Home Office (Grant No RDS/01/247) and the NHS Forensic Mental Health Research and Development Programme (Grant No. MRD/60).

Notes

1. There are further variants of 4-factor models which have been empirically tested, which are discussed elsewhere (see Cooke et al., 2007).
2. This is in line with the International Classification of Diseases-10 (ICD-10; World Health Organisation, 1992) and *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition, Text Revision (American Psychiatric Association, 2000) criteria for mental retardation. The term is synonymous with that of "learning disabilities" in the United Kingdom and "developmental delay" or "mental retardation" in North America.
3. Most participants were classified as having mild disabilities. The criteria for determining Mental Retardation in the main diagnostic systems is a Full Scale IQ of less than 70. However there were nine individuals included in this study with IQs between 70 and 75, allowing some flexibility for confidence intervals for Wechsler Adult Intelligence Scale–III scores. These individuals would have had significant deficits in adaptive functioning, and had been assessed as requiring intellectual disability services.
4. Because of the identified difference in the age of the samples, all analysis was repeated using similarly sized age-matched samples. No differences in the findings of the confirmatory factor analysis or item response theory analyses were observed. These analyses are available from the first author on request.
5. The original version of these guidelines was approved for use for research purposes by R. D. Hare (personal communication, February 2003). The updated guidelines are available at http://institutemh.org.uk/uploads/upload_5249.
6. Although only the 3-factor model fits the data, it will be noted that IRT analyses were conducted for both the 13- and 20-item tests, so that item parameters for both versions of the test could be examined. The fact that the results for the 13 items are similar for the two sets of results suggests that the 20 items do form a sufficiently unidimensional test to justify reporting these data.

References

Alexander, R., & Cooray, S. (2003). Diagnosis of personality disorders in learning disability. *British Journal of Psychiatry*, *44*(Suppl.), 528-531.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., Text Rev.). Washington DC: Author.

Bentler, P. M., & Wu, E. J. C. (1995). EQS for Windows [Computer software]. Encino, CA: Multivariate Software.

Bolt, D. M., Hare, R. D., & Vitale, J. E. (2004). A multigroup item response theory analysis of the Psychopathy Checklist-Revised. *Psychological Assessment*, *16*, 155-168.

Cleckley, H. (1941). *The mask of sanity.* St Louis, MO: Mosby.

Cleckley, H. (1976). *The Mask of Sanity* (5th ed.). St Louis, MO: Mosby.

Cooke, D. J., Kosson, D. S., & Michie, C. (2001). Psychopathy and ethnicity: Structural, item and test generalizability of the Psychopathy Checklist–Revised (PCL-R) in Caucasian and African American participants. *Psychological Assessment*, *13*, 531-542.

Cooke, D. J., & Michie, C. (1997). An item response theory analysis of the Hare Psychopathy Checklist. *Psychological Assessment*, *9*(3), 3-13.

Cooke, D. J., & Michie, C. (1999). Psychopathy across cultures: Scotland and North America compared. *Journal of Abnormal Psychology*, *108*, 58-68.

Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment*, *13*, 171-188.

Cooke, D. J., & Michie, C. (2009). Limitations of diagnostic precision and predictive utility in the individual case: A challenge for forensic practice. *Law and Human Behavior.* Retrieved August 11, 2009, from https://commerce.metapress.com/content/n1463680m4167252/

Cooke, D. J., Michie, C., Hart, S. D., & Clark, D. A. (2004). Reconstructing psychopathy: Clarifying the significance of antisocial and socially deviant behavior in the diagnosis of psychopathic personality disorder. *Journal of Personality Disorders, 18*, 337-356.

Cooke, D. J., Michie, C., Hart, S. D., & Clark, D. A. (2005a). Assessing psychopathy in the UK: Concerns about cross-cultural generalisability. *British Journal of Psychiatry*, *186*, 335-341.

Cooke, D. J., Michie, C., Hart, S. D., & Clark, D. (2005b). Searching for the pan-cultural core of psychopathic personality disorder. *Personality and Individual Differences*, *39*, 283-295.

Cooke, D. J., Michie, C., & Skeem, J. (2007). Understanding the structure of the Psychopathy Checklist–Revised: An exploration of methodological confusion. *British Journal of Psychiatry*, *190*, s3-s7.

Department of Health. (2009). *Lord Bradley's review of people with mental health problems or learning disabilities in the criminal justice system.* London: Author.

Douglas, K. S., Yeomans, M., & Boer, D. (2005). Comparative validity of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior*, *32*, 479-510.

Edens, J. F., Petrila, J., & Buffington-Vollum, J. K. (2001). Psychopathy and the death penalty: Can the Psychopathy Checklist–Revised identify offenders who represent "an ongoing threat to society?" *Journal of Psychiatry & Law*, *29*, 433-481.

Embretson, S., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Fazel, S., Xenitidis, K., & Powell, J. (2008). The prevalence of intellectual disabilities among among 12,000 prisoners: A systematic review. *International Journal of Law and Psychiatry*, *31*, 369-373.

Forouzan, E., & Cooke, D. J. (2005). Figuring out la femme fatale: Conceptual and assessment issues concerning psychopathy in females. *Behavioral Sciences & the Law*, *23*, 765-778.

Forth, A. E., Kosson, D. & Hare, R. D. (2003). *Hare Psychopathy Checklist: Youth Version* (PCL: YV). Toronto, Ontario: Multi Health Systems.

Gray, N., Fitzgerald, S., Taylor, J., MacCulloch, M. J., & Snowden, R. J., (2007). Predicting future reconviction in offenders with intellectual disabilities: The predictive efficacy of VRAG, PCL-SV, and the HCR-20. *Psychological Assessment*, *19*, 474-479.

Guy, L. S., Edens, J. F., Anthony, C., & Douglas, K. S. (2005). Does psychopathy predict institutional misconduct among adults? A meta-analytic investigation. *Journal of Consulting and Clinical Psychology*, *73*, 1056-1064.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage.

Hare, R. D. (1991). *The Hare Psychopathy Checklist–Revised*. Toronto, Ontario, Canada: Multi-Health Systems.

Hare, R. D. (2003). *The Hare Psychopathy Checklist–Revised* (2nd ed.). Toronto, Ontario, Canada: Multi-Health Systems.

Hare, R. D., & Neumann, C. S. (2005). Structural models of psychopathy. *Current Psychiatry Reports, 7*, 57-64.

Harpur, T. J., Hakistan, R., & Hare, R. D. (1988). Factor structure of the Psychopathy Checklist. *Journal of Consulting and Clinical Psychology*, *56*, 741-747.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *The Hare PCL:SV*. Toronto, Ontario: Multi-Health Systems.

Hemphill, J. F., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology*, *3*, 139-170.

Hogue, T., Steptoe, L., Taylor, J., Lindsay, W., Mooney, P., Pinkney, L., et al., (2006). A comparison of offenders with intellectual disability across three levels of security. *Criminal Behavior and Mental Health*, *16*, 16-28.

Holland, T., **Clare**, I. C. H., & Mukhopadhyay, T. (2002). Prevalence of criminal offending by men and women with intellectual disability and characteristics of offenders: Implications for research and service development. *Journal of Intellectual Disability Research*, *46*(Suppl. 1), 6-20.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York; Guilford.

Kosson, D. S., Smith, S. S., & Newman, J. P. (1990). Evaluating the construct validity of psychopathy in black and white male inmates: Three preliminary studies. *Journal of Abnormal Psychology*, *99*, 250-259.

Lilienfield, S. (1994). Conceptual problems in the assessment of psychopathy. *Clinical Psychology Review*, *14*, 17-38.

Lindsay, W. R., Hogue, T., Taylor, J. L., Mooney, P., Steptoe, L., Johnston, S. O., et al. (2006). Two studies of the prevalence and validity of personality disorder in three forensic intellectual disability samples. *Journal of Forensic Psychiatry & Psychology*, *17*, 485-506.

Lyall, I., Holland, A. J., & Collins S. (1995). Offending by adults with learning disabilities: Identifying need in one health district. *Mental Handicap Research*, *8*, 99-109.

McBrien, J., & Murphy, G. (2006). Police and carers' views on reporting alleged offenders by people with intellectual disabilities. *Psychology, Crime & Law*, *12*, 127-144.

McCrae, R. R., & Costa, P. T., Jr., (2003). *Personality in adulthood: A five factor theory perspective* (2nd ed.). New York: Guilford.

Morrissey, C. (2005). *Guidelines for assessing psychopathy in offenders with intellectual disabilities using the PCL:R and PCL:SV*. Retrieved July 9, 2009, from http://institutemh.org.uk/uploads/upload_5249.pdf

Morrissey, C., Hogue, T., Mooney, P., Allen, C., & Johnston, S., Hollin, C., et al., (2007a). The predictive validity of the PCL-R in offenders with intellectual disability in a high secure hospital setting: Institutional aggression. *Journal of Forensic Psychiatry & Psychology*, *18*, 1-15.

Morrissey, C., Hogue, T., Mooney, P., Lindsay, W. R., Steptoe, L., Taylor, J., et al., (2005). Applicability, reliability, and validity of the Psychopathy Checklist–Revised in offenders with intellectual disabilities: Some initial findings. *International Journal of Forensic Mental Health*, *4*, 207-220.

Morrissey, C., Mooney, P., Hogue, T., Lindsay, W. R., & Taylor, J. (2007b). Predictive validity of psychopathy in offenders with intellectual disability in high security: Treatment progress. *Journal of Intellectual and Developmental Disability*, *32*, 125-134.

Murphy, G. H. (1992). Community adjusted social integration, work and social competence for people with intellectual disabilities or mental retardation. *Current Opinion in Psychiatry, 5,* 831-835.

Raju, N., Linden, W., & Fleer, P. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, *19*, 353-368.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psvchometrika Monograph*, No. 17.

Satorra, A., & Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514.

Skeem, J., & Cooke, D. J. (in press). Is criminal behavior a central component of psychopathy? Conceptual directions for resolving the debate. *Psychological Assessment.*

Thissen, D. (2003). *Multilog user's guide* (Version 7). Mooresville, NJ: Scientific Software International.

Vitacco, M. J. (2007). Psychopathy and antisocial tendencies: A reply to Cooke, Michie and Skeem [Letter to the editor]. *British Journal of Psychiatry*, *191*, 357.

Vitale, J. E., Smith, S. S., Brinkley, C. A., & Newman, J. P. (2002). The reliability and validity of the Psychopathy Checklist-Revised in a sample of female offenders. *Criminal Justice and Behavior*, *29*, 202-231.

Walters, G. D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. *Law and Human Behavior*, *27*, 541-558.

World Health Organisation. (1992). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: Author.